

Міністерство освіти і науки України Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського" Фізико-технічний інститут

> "КРИПТОГРАФІЯ" Комп'ютерний практикум Робота № 1

Виконав:	
	студент групи ФБ-93 Проценко О.А.
Перевірила:	

Селюх П. В.

1. Назва роботи

Експериментальна оцінка ентропії на символ джерела відкритого тексту.

Мета

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

2. Особливості роботи:

- Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
- Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку 1 Н та 2 Н за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення 1 Н та 2 Н на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення 1 Н та 2 Н на тому ж тексті, в якому вилучено всі пробіли.
- За допомогою програми CoolPinkProgram оцінити значення $H^{(10)}$, $H^{(20)}$, $H^{(30)}$.
- Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.
- 3. Робота виконана 06.10.2021 12.10.2021 року в місті Київ. Робота створена на мові програмування Руthon.
- 4. Робота над завданням
 - ▶ Програма, яка підраховує частоти букв та біграм (що перетинаються і не перетинаються). Для визначення кількості букв (пробілу) використав звичайну посимвольну акумуляцію. Для визначення кількості біграм використав контейнер "collections" а саме об'єкт "Counter" в парі з функцією "range(strart, stop, step)" де і зазначав крок дій. В результаті дані використовую для підрахунку їх частотності.

Складність полягала в недостачі знань роботи в мові програмування. Все було вирішено, прочитавши потрібні інструкції до команд та саму документацію щодо лабораторної.

Частота букв без пробілу

- o >>> 0.11275224397101763 time(s)
- e >>> 0.08453768789877798 time(s)
- a >>> 0.07721855736995782 time(s)
- $_{\rm H} >>> 0.06976316643235644$ time(s)
- H >>> 0.0640250892181248 time(s)
- T >>> 0.060097328863415164 time(s)
- c >>> 0.050762409430085434 time(s)
- B >>> 0.04583324321401536 time(s)
- $\pi >>> 0.04576186871417757$ time(s)
- p >>> 0.042398615767275875 time(s)
- $\kappa >>> 0.04061641613496269 \text{ time(s)}$
- $\pi >>> 0.03115388774737753 \text{ time(s)}$
- M >>> 0.03007245593165351 time(s)
- y >>> 0.029680977614361413 time(s)
- $\pi >>> 0.028376770844598248$ time(s)
- ь >>> 0.019610684546339352 time(s)

```
ы >>> 0.018968314047799285 time(s)
```

$$g >>> 0.018596301503190225$$
 time(s)

$$\Gamma >>> 0.017633827187195847$$
 time(s)

$$\kappa >>> 0.010822969611765978$$
 time(s)

$$\text{III} >>> 0.009771817886882233 \text{ time(s)}$$

$$\mu >>> 0.003611982264518222 \text{ time(s)}$$

$$_{\text{III}} >>> 0.0031664323564399265 \text{ time(s)}$$

$$\gamma >>> 0.0027100681302043906$$
 time(s)

$$\phi >>> 0.0015377960419595544 \text{ time(s)}$$

$$_{\rm b} >>> 0.0003071266356656213$$
 time(s)

$$\ddot{e} >>> 0.00010165459067805775 \text{ time(s)}$$

Ентропія букв без пробілу >>> 4.472830206355587

$$R = 1 - \frac{4.472}{5} = 0.1056$$
 10.56%

Частота букв з пробілом

- _>>> 0.1611100527084524 time(s)
- o >>> 0.0945867240018507 time(s)
- e >>> 0.07091781654555517 time(s)
- a >>> 0.06477787152201327 time(s)
- $_{\rm H} >>> 0.058523619011330956$ time(s)
- $_{\rm H} >>> 0.053710003719529345 \, {\rm time(s)}$
- T >>> 0.05041504504259315 time(s)
- c >>> 0.042584074971196324 time(s)
- B >>> 0.03844904698400602 time(s)
- $\pi >>> 0.038389171633599145$ time(s)
- p >>> 0.03556777254624464 time(s)
- $\kappa >>> 0.03407270319063042 \text{ time(s)}$
- $_{\rm I} >>> 0.026134683250324326$ time(s)
- M >>> 0.0252274809714322 time(s)
- y >>> 0.024899073746473252 time(s)
- $\pi >>> 0.02380498779812935$ time(s)
- ь >>> 0.016451206125429788 time(s)
- ы >>> 0.015912327971767865 time(s)
- g >>> 0.015600250387828975time(s)
- ч >>> 0.015516787778170898 time(s)
- 6 >>> 0.015362563390759238 time(s)
- $\Gamma >>> 0.014792840359614983$ time(s)
- 3 >>> 0.014243075778606356 time(s)
- $\kappa >>> 0.009079280407152383 \text{ time(s)}$
- й >>> 0.00876175960954014 time(s)

III >>> 0.008197479792069237 time(s)

x >>> 0.008008781718059675time(s)

ю >>> 0.005310762140634497 time(s)

ц >>> 0.003030055611499696 time(s)

 $_{\rm III} >>> 0.0026562882725961407 \, {\rm time(s)}$

э >>> 0.002273448910903664 time(s)

φ >>> 0.0012900416405846012 time(s)

 $_{\rm b} >>> 0.0002576454472053634 \, {\rm time(s)}$

ë >>> 8.527701421585971e-05 time(s)

Ентропія букв з пробілом >>> 4.389170847948409

$$R = 1 - \frac{4.389}{5} = 0.1222$$
 12.22%

Біграма з пробілом 1 крок (перші 20)

o >>> 0.021331993090734906 time(s)

и >>> 0.01835273540127444 time(s)

e >>> 0.018309189612877943 time(s)

 $\pi >>> 0.017079021090676848$ time(s)

H >>> 0.016237135848344533 time(s)

B >>> 0.015230139491675497 time(s)

c >>> 0.015226510675975788 time(s)

a >>> 0.014711218846617217 time(s)

a_ >>> 0.014/1121884001/21/time(s)

то >>> 0.01296938731075726 time(s)

 $_{\text{U}} >>> 0.0102350746810271 \text{ time(s)}$

ко >>> 0.010193343300480455 time(s)

cT >>> 0.010069963566690375 time(s)

ь >>> 0.01004093304109271 time(s)

 π o >>> 0.009785101534263279 time(s)

но >>> 0.009652649761223927 time(s)

 $\kappa >>> 0.009612732788527137 \text{ time(s)}$

на >>> 0.009307912269751644 time(s)

o >>> 0.009258923257805583 time(s)

я >>> 0.00913010030046594 time(s)

не >>> 0.009033936684423672 time(s)

Ентропія біграми з пробілом 1 крок >>> 3.97403625359661

$$R = 1 - \frac{3.974}{5} = 0.2052$$
 20.52%

Біграма з пробілом 2 кроки (перші 20)

o >>> 0.021333807498584764 time(s)

и >>> 0.018234798891033924 time(s)

e >>> 0.01815496494564034 time(s)

 $\pi >>> 0.017182442338118532$ time(s)

 $_{\rm H} >>> 0.016333299464386802$ time(s)

B >>> 0.015560361720348947 time(s)

c >>> 0.015404322645261492 time(s)

```
a >>> 0.014489861088935015 time(s)
```

$$\pi$$
o >>> 0.009866749887506713 time(s)

$$n_{>>>} 0.009075668064970316$$
time(s)

Ентропія біграми з пробілом 2 кроки >>> 3.9730008821084835

$$R = 1 - \frac{3.973}{5} = 0.2054$$
 20.54%

Біграма без пробілу 1 крок (перші 20)

$$c_T >>> 0.01231104641731679 \text{ time(s)}$$

He
$$>>> 0.010842458835208901$$
 time(s)

$$\kappa a >>> 0.01019143547406829 \text{ time(s)}$$

$$a\kappa >>> 0.009198678920036596 time(s)$$

$$o_T >>> 0.008854782858836074 \text{ time(s)}$$

$$a\pi >>> 0.008824502702503952 \text{ time(s)}$$

$$\pi p >>> 0.008539004085658237 \text{ time(s)}$$

$$\pi u >>> 0.008247016863884208 \text{ time(s)}$$

Ентропія біграми без пробілу 1 крок >>> 4.143481201154285

$$R = 1 - \frac{4.143}{5} = 0.1714$$
 17.14%

Біграма без пробілу 2 кроки (перші 20)

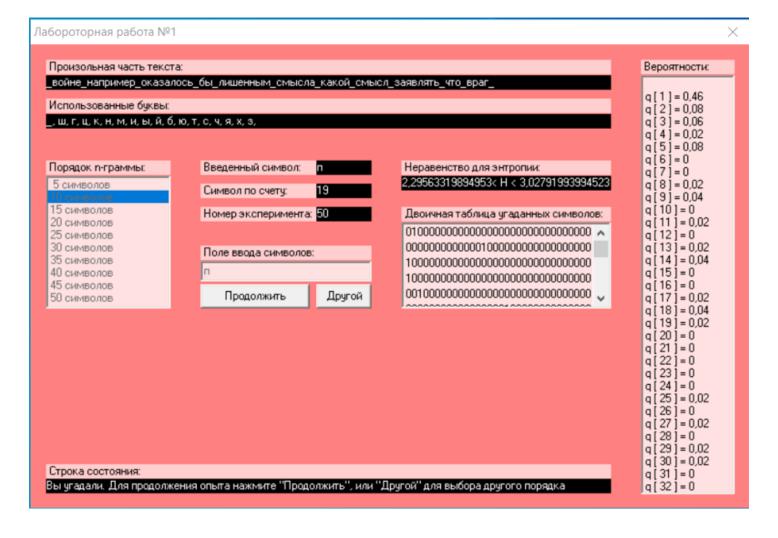
$$c_T >>> 0.012821455607223965 \text{ time(s)}$$

```
ко >>> 0.012233156699470098 time(s)
но >>> 0.011982264518222126 time(s)
\pi o >>> 0.011489131610251974 \text{ time(s)}
на >>> 0.011099816156591327 time(s)
не >>> 0.010840272520817562 time(s)
ка >>> 0.01010489888612523 time(s)
ен >>> 0.009892938250243323 time(s)
a\kappa >>> 0.009529577160160053 \text{ time(s)}
_{\rm H II} >>> 0.009473342705742403 \, {\rm time}({\rm s})
ro >>> 0.009443062614902131 time(s)
oc >>> 0.009066724343030172 time(s)
ot >>> 0.009053747161241485 time(s)
po >>> 0.008962906888720666 time(s)
он >>> 0.00887206661619985 time(s)
ал >>> 0.00871634043473559 time(s)
\pi p >>> 0.008586568616848709 \text{ time(s)}
Bo >>> 0.007993943981831945 time(s)
```

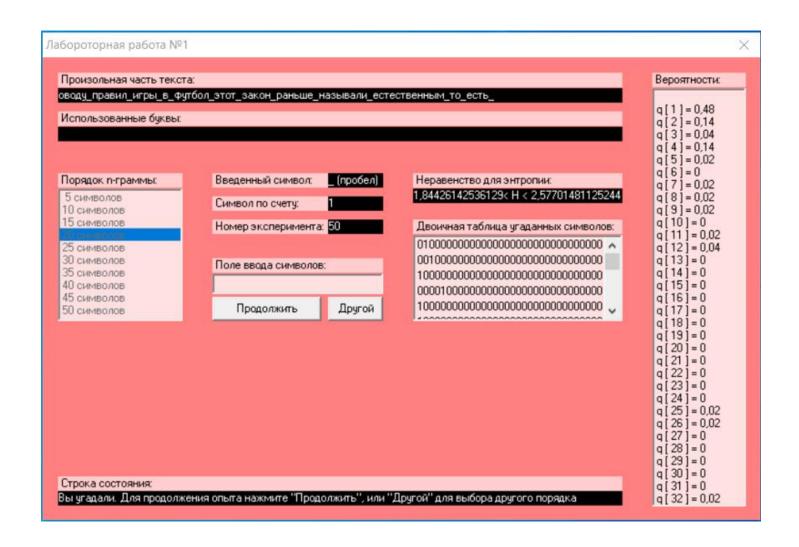
Ентропія біграми без пробілу 2 кроки >>> 4.142553385816108

$$R = 1 - \frac{4.142}{5} = 0.1716$$
 17.16%

CoolPinkProgram

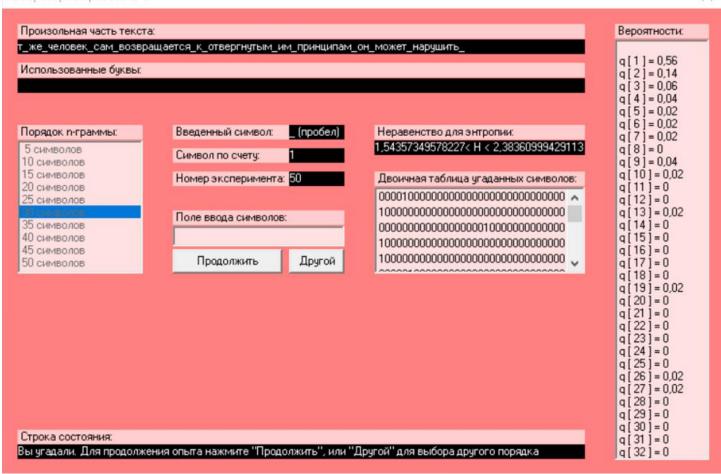


$$2.295 < H < 3.027$$
 $H^{(10)} = \frac{2.295 + 3.027}{2} = 2.661$



$$1.844 < H < 2.577$$

$$H^{(20)} = \frac{1.844 + 2.577}{2} = 2.2105$$



$$1.543 < H < 2.383$$

$$H^{(30)} = \frac{1.543 + 2.383}{2} = 1.963$$

> Оцінка Надлишковості

$$R_{10} = 1 - \frac{2.661}{5} = 0.4678 46.78\%$$

$$R_{20} = 1 - \frac{2.2105}{5} = 0.5579$$
 55.79%

$$R_{30} = 1 - \frac{1.963}{5} = 0.6074 \tag{60.74\%}$$

5. Висновки.

Отримав навички роботи з мовою програмування Руthon. Дізнався як обчислити ентропію, частотність та надлишковість. Розібрався з дослідженням в цілому. Під час дослідження роботи було визначено що найвживаніші букви в російському алфавіті — '_'; 'o'; 'e'; 'a'; щодо менш вживаних — 'ë'; 'ь'; 'ф'; 'э'. Найвживаніші біграми (з пробілом) в російській мові — 'o_'; 'и_'; 'e_'; '_п'; менш вживані (без пробілу) — 'пз'; 'фф'; 'фф'; 'бф'. Також, використавши програму CoolPinkProgram, встановив що середня надлишковвсть російської мови становить 54.43%.